MBMG
Metabarcoding & Metagenomics

Nanopore duplex sequencing as an alternative to Illumina MiSeq sequencing for eDNA-based biomonitoring of coastal aquaculture impacts

Thorsten Stoeck[1], Sven Nicolai Katzenmeier[1], Hans-Werner Breiner[1], Verena Rubel[1]

1   *Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau, Ecology Group, Kaiserslautern, Germany*
Corresponding author: Thorsten Stoeck (stoeck@rptu.de)

OPEN ACCESS

Citation: Stoeck T, Katzenmeier SN, Breiner H-W, Rubel V (2024) Nanopore duplex sequencing as an alternative to Illumina MiSeq sequencing for eDNA-based biomonitoring of coastal aquaculture impacts. Metabarcoding and Metagenomics 8: e121817. https://doi.org/10.3897/mbmg.8.121817

Abstract

Oxford Nanopore Technologies has recently launched a duplex sequencing strategy and announced an improved error rate which is in a similar order of magnitude as Illumina sequencing. We therefore conducted a pilot study to assess whether Nanopore duplex sequencing has potential to be used as a technology in eDNA-based marine biomonitoring. Specifically, we investigated bacterial communities of sediment samples collected from Atlantic salmon (*Salmo salar*) aquaculture installations and compared the ecological trends obtained from short Illumina (V3-V4 region of the 16S rRNA gene) and long Nanopore (full length 16S rRNA gene) sequence reads. The obtained duplex rate of Nanopore amplicon reads with a Phred score ≥ 30 was 36%, notably higher compared to previous reports from bacterial genome sequencing. When inferring alpha- and beta-diversity from Illumina ASVs and Nanopore OTUs, we found highly congruent ecological patterns. Only when collating ASVs and OTUs across taxonomic ranks, beta-diversity analyses of Illumina-data slightly changed, due to the difficulties to assign a taxonomy to short sequence reads. While on family rank, both sequence datasets had good agreement, genus-assignments of Illumina data were critical, resulting in higher disagreement between the two protocols. Our data provide evidence that eDNA-based monitoring of aquaculture-related environmental impacts could equally well be conducted with the improved Nanopore duplex sequencing. We discuss to what extent eDNA-based biomonitoring could benefit from long-read information.

Key words: Aquaculture impacts, eDNA-based biomonitoring, metabarcoding, microbial communities, Nanopore sequencing, 16S rRNA gene

Introduction

Traditional marine ecosystem biomonitoring in research and in practice (compliance monitoring) relies on the morphology-based identification of macroinvertebrate indicators. This is expensive, requires a high-level of taxonomic expertise and, above all, takes more time to produce results than is needed for active management (Cordier et al. 2019). In addition, this approach cannot be easily up-scaled to meet increasing sample numbers in updated compliance monitoring regulations (see, for example, SEPA (2023)). Therefore, DNA-based technologies have been

explored with success and are increasingly replacing macroinvertebrate-based biomonitoring protocols. For marine ecosystems, especially bacterial communities have emerged as informative and powerful bioindicators for, for example, biomonitoring the effects of aquaculture installations (Dowle et al. 2015; Keeley et al. 2018; Stoeck et al. 2018; Dully et al. 2021a; Frühe et al. 2021a; Leontidou et al. 2023; Wilding et al. 2023), oil- and gas-extraction platforms (Laroche et al. 2018; Oladi et al. 2022), marine construction (Aylagas et al. 2017) and pollutant discharge from land-based industries (Aylagas et al. 2017; Borja 2018; Lanzén et al. 2021) on marine ecosystem health. The general approach for the analysis of natural microbial communities is the extraction of DNA from environmental samples (eDNA) and the PCR amplification of the V3-V4 hypervariable region of the bacterial 16S ribosomal RNA (16S rRNA) as the bacterial marker gene using high-throughput sequencing. This is then followed by taxonomic classification of the obtained amplicons. Illumina sequencing is the dominating technology in environmental sequencing of microbial communities. Illumina sequencing, a so-called second-generation (or next generation) sequencing technology, is highly accurate, extremely powerful and successful in microbial ecology. However, due to the maximum amplicon size of ca. 500 bp (MiSeq platform, Ravi et al. (2018)) Illumina sequencing has a limited capacity regarding the taxonomic assignment of these short-read amplicons compared to the full-length 16S rRNA genes: in general, short-read Illumina amplicons, such as the V3-V4 16S rRNA gene, cannot be reliably assigned to lower taxonomic levels (genus and species ranks) (Older et al. 2023; Petrone et al. 2023; Zhang et al. 2023).

A reliable identification of bacterial species, based on 16S rRNA genes, requires the (nearly) complete sequence information included in this taxonomic marker gene (Johnson et al. 2019; Petrone et al. 2023). Therefore, specific molecular bioindicators in current biomonitoring protocols that were inferred from short-read Illumina datasets are typically so-called ASVs (amplicon sequence variants), often without a taxonomic framework. Even though ASV-based bioindication does not necessarily require a taxonomic assignment of the indicator ASVs (Cordier et al. 2019), an accurate species assignment of these indicators would provide additional ecological information that could be of high relevance for bioindication. A further benefit of long-range reads, especially in biomonitoring, is that substantially more information is available for the design of indicator-specific PCR primers to develop sequencing-free, fast-screening bioindicator assays using quantitative PCR approaches, such as digital PCR (Netzer et al. 2021). Furthermore, Illumina sequencing requires expensive hardware, which is rarely affordable for individual laboratories, resulting as a rule in an outsourcing of Illumina sequencing to commercial providers. This sometimes comes with longer waiting times to obtain results and makes own optimisations of sequencing protocols for specific sample material or applications practically impossible.

Therefore, optimisation of the sequencing step in eDNA-based biomonitoring protocols would be the integration of a long-range technology that can compete with Illumina sequencing in accuracy and massive data production. One possibility that has been explored for taxonomic profiling of microbial communities was Nanopore sequencing (e.g. Nygaard et al. (2020); Karst et al. (2021); Girija et al. (2023); Klair et al. (2023); Lemoinne et al. (2023); Older et al. (2023); Petrone et al. (2023); Stevens et al. (2023); Tandon et al. (2023); Zhang et al. (2023); Zorz et al. (2023)), a third-generation sequencing technology, developed by Oxford Nanopore

Technologies (ONT) (for a detailed description of Nanopore sequencing, see, for example, Wang et al. (2021) and MacKenzie and Argyropoulos (2023)). However, until recently, Nanopore-derived sequences were not suitable for biomonitoring applications, which requires an as exact as possible sequence accuracy (Cordier et al. 2019; Cordier et al. 2021) and an as high as possible signal-to-noise ratio (Wilding et al. 2023) for the identification of individual bioindicator ASVs. While the median error rate of Illumina sequencing (MiSeq) is about 0.5% (Goodwin et al. 2016; Stoler and Nekrutenko 2021), the error rates reported for data generated with Nanopore sequencing on an R9.4.1 flow cell were notably higher. Typical error rates reported were for example 6% (Delahaye and Nicolas 2021), 8.8% (Zhang et al. 2023) and even up to 15% (Wang et al. 2021). With the introduction of the sequencing kit LSK112 Q20+ (> 99% accuracy) and the new flow cell R10.4, error rates dropped down to ca. 2- 4% in bacterial 16S rRNA profiling (Zhang et al. 2023).

Very recently, a further advancement in Nanopore sequencing was duplex sequencing, for which ONT reported an accuracy of > 99.9% (Q30) in combination with super-accuracy base-calling models and a duplex rate of > 70% when sequencing *Escherichia coli* (ONT 2024). Duplex sequencing, first described as a concept more than a decade ago (Chen et al. 2010), is accomplished in Nanopore sequencing by anchoring one DNA strand in place with a hairpin adapter, waiting for the complete translocation of the other strand through the nanopore before threading through the sister strand (MacKenzie and Argyropoulos 2023). By sequencing both strands of a DNA molecule, errors that occur in one strand can be identified and corrected using information from the complementary strand. This redundancy helps to reduce sequencing errors. Furthermore, duplex reads can enhance the confidence in base-calling for each position in the DNA sequence. If the two strands agree on a particular base, it provides additional support for the accuracy of that base-call. Additionally, duplex sequencing may improve the discrimination of homopolymers, which used to be a known bias in Nanopore sequencing (Delahaye and Nicolas 2021). These improvements could make Nanopore sequencing an attractive option for the implementation in compliance biomonitoring protocols using bacterial communities as bioindicators.

To the best of our knowledge, no peer-reviewed publication has taken ONT's Q30 sequencing chemistry to the test. However, prior research employing Q20+ chemistry has already conducted a comparison between the Illumina MiSeq platform and ONT sequencing. It revealed that the longer sequencing lengths can offset diminished sequence quality, enabling a comparable identification of bacterial communities at higher taxonomic levels between the two platforms (Older et al. 2023; Petrone et al. 2023; Zhang et al. 2023). Most of these studies have analysed standardised bacterial mock communities, with limited relevance for biomonitoring and aquaculture (Older et al. 2023). While certain findings may be applicable to other sample types and organisms, the most effective methodologies should be selected according to the specific habitat and type of sample under consideration.

In this pilot study, we therefore analysed samples collected along an organic enrichment gradient from two salmon farms in Scotland. Using the same eDNA extracts from these samples, we analysed the bacterial community structures with both Illumina MiSeq and the new Nanopore Q30 chemistry with duplex reads. The aim of this proof-of-principle study was to compare measures of alpha- and beta-diversity in natural fish-farming related bacterial communities which are of relevance in biomonitoring.

Material and methods

Sample collection

Samples were collected from two salmon farm locations in Scotland, namely DUN located close to Oban and LIS located in Loch Linnhe. Sampling was conducted during the mid-production and peak-production period, respectively. Sediment was collected at three sites along a transect from an outer cage edge (CE) to a reference site (REF) in the direction of the prevailing current flow, located ca. 800 m (DUN) and 525 m (LIS) distant from the CE. An intermediate impact zone (Allowable Zone of Effect, AZE) was located at ca. 100 m (DUN) and 109 m (LIS) distance from the cage edge. This sampling design followed a decreasing organic enrichment gradient from the CE towards the REF sites, resulting from the deposits of feed and fish faeces on the sea floor (Keeley et al. 2012; Bannister et al. 2014; Frühe et al. 2021b). For a graphical representation of the sampling design, we refer to Fig. 1. At each site, two biological replicates were taken with a van Veen grab (0.1 m^2 area, DUN; 0.045 m^2 area, LIS). From each replicate, we sampled approximately 20–25 g of surface sediment (upper few millimetres) into a sterile 50 ml plastic tube using disposable sterile plastic spatulas. Immediately following collection, we homogenised the samples in the 50-ml collection tube with a spatula. Samples were frozen within a few hours of collection and shipped (< 48 h) from Scotland to our laboratory in Kaiserslautern (Germany) on dry-ice. Samples from farm LIS were taken during routine compliance monitoring and are accompanied by macroinvertebrate inventories which were used to calculate the AMBI Index and ecological quality (Borja et al. 2000; Muxika et al. 2005). The macrofauna was obtained from van Veen grabs after subsampling for DNA analyses. Therefore, the remaining sediment was washed through a 1 mm sieve and the residue was fixed in 4% borax-buffered formaldehyde prior to macrobenthic sorting and counting. The sieve-retained fauna was identified to species level under the National Marine Biological Quality Control Scheme (NMBAQCS) by APEM Ltd., Hertfordshire.

DNA extraction and PCR amplifications

Following our previously described protocol (Frühe et al. 2021a), environmental DNA was obtained from sediment samples using the PowerSoil DNA Kit (Qiagen, Hilden, Germany) according to the manufacturer's manual. The concentration and quality of extracted DNA were measured with a NanoDrop 2000 spectrophotometer (Peqlab, Erlangen, Germany).

As DNA metabarcodes for Illumina MiSeq sequencing, we obtained the ca. 450 bp long hypervariable V3-V4 region of the bacterial 16S rRNA gene using the Bact_341F (5'-CCTACGGGNGGCWGCAG-3') and the Bact_805R (5'-GACTACHVGGGTATCTAATCC-3') primer pair (Herlemann et al. 2011). PCR conditions employed an initial activation step of NEB's Phusion High-Fidelity DNA polymerase at 98 °C for 30 s, followed by 27 identical three-step cycles consisting of 98 °C for 10 s, 62 °C for 30 s and 72 °C for 30 s; then a final 5-min extension at 72 °C.

For Nanopore sequencing, we obtained the full length 16S rRNA gene using primer pair Bac27f (5'-AGAGTTTGATCMTGGCTCAG-3') (Frank et al. 2008) and
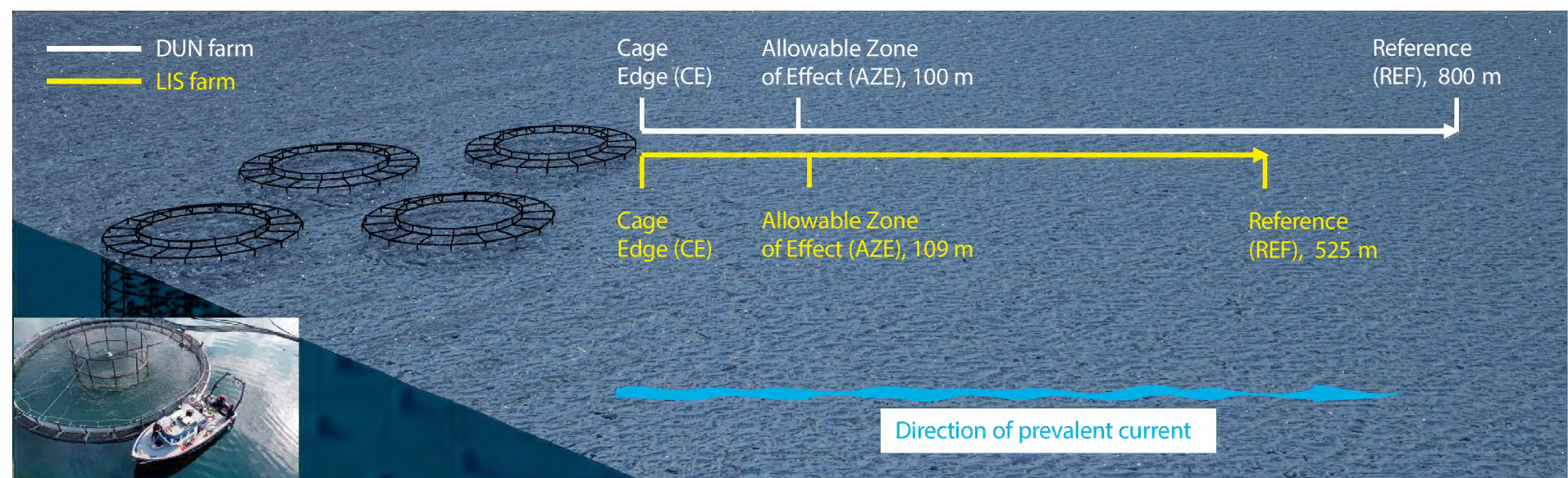
Figure 1. Schematic representation of the sampling design. Benthic sampling of the two salmon aquaculture installations (DUN and LIS) occurred along a transect from the outermost cage edge of each farm, towards a reference site (REF). Distances of the Allowable Zone of Effect (AZE) and the REF sampling sites are determined according to compliance monitoring regulations separately for each farm prior to stocking. These distances depend on different parameters, such as strength of current and seabed topology. Distances of reference sites are chosen to be outside the influence of seabed depositions originating from aquaculture activities. The inset image (lower left corner) shows the RV Seol Mara of SAMS, Oban, while sampling the CE of DUN farm.

U1492R (5'- ACCTTGTTACGRCTT-3') (Dawson and Pace 2002), both of which are more universal modifications of the bacterial 27f/1492R primer pair developed by Lane (1991). For amplification, we used the LongAmp Taq polymerase (New England Biolabs, Ipswich, MA, USA). The PCR protocol consisted of an initial activation step at 95 °C for 60 s, followed by 30 identical three-step cycles consisting of 95 °C for 20 s, 55 °C for 30 s and 65 °C for 2 min; then a final 5-min extension at 65 °C.

In both cases, V3-V4 region and full-length 16S rRNA gene, we obtained three technical PCR replicates for each of the two grab replicates per sampling site. Resulting PCR products were then purified using Qiagen's MinElute PCR purification kit. All six PCR replicates of one sampling site (2 sediment grabs × 3 technical PCR replicates per grab) were then pooled in equal amounts of DNA prior to library constructions.

Library constructions and sequencing

Illumina: From the resulting PCR products, sequencing libraries were constructed using the NEB Next® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA). The quality of the libraries was assessed with an Agilent Bioanalyzer 2100 system. V3-V4 libraries were sequenced on an Illumina MiSeq platform, generating 2 × 300-bp paired-end reads. Illumina short-read amplicons are available in NCBI's BioProject database under BioProject ID PRJNA768445.

Nanopore: Libraries for Nanopore sequencing were constructed using the Duplex-enabled Native Barcoding Kit 24 V14 (SQK-NBD114.24) (Oxford Nanopore Technologies, ONT), following the manufacturer's manual. Sequencing of libraries (15 ng DNA total, measured with a Quantus Fluorometer (Promega) was performed on a MinION sequencing device (MIN-101B, ONT) and an R10.4.1 flow cell (FLO-MIN114, ONT). Nanopore sequence reads were deposited in NCBI's BioProject database under BioProject ID PRJNA1076812.

Sequence reads processing

Sequence data processing of Illumina R1 and R2 output files and of Nanopore pod5 output files were conducted on the high-performance computing cluster (HPC) of the RPTU Kaiserslautern-Landau.

Illumina: Raw sequence reads were quality filtered and trimmed by executing the dada2 (divisive amplicon denoising algorithm) workflow (Callahan et al. 2016) in R Studio 3.5.1 to obtain ASVs (Amplicon Sequence Variants). Truncation length was set to 255 bp (Dully et al. 2021b) and we kept only sequences with a mean Phred quality score ≥ 30 (Q3) corresponding to 99.9% base-call accuracy (Ewing and Green 1998). For maxEE, we chose 1 to maximize downstream sequence quality. The paired-end sequences were merged using a minimum 20 bp overlap and a mismatch of two bases was allowed (Frühe et al. 2021a). Before the construction of the ASV-to-sample matrix, the sequences were checked for chimeras using the *uchime_denovo* function of *vsearch* (Rognes et al. 2016). ASVs that were represented by only one single sequence in the complete Illumina dataset (singletons) were removed from the final ASV-to-sample matrix.

Nanopore: Pod5 files were subjected to duplex base calling using ONT's Dorado duplex base-caller version 0.5.3 with the super-accuracy model (dna_r10.4.1_e8.2_400bps_sup@v4.2.0, https://github.com/nanoporetech/dorado) (Wick 2023). SAMtools (Li et al. 2009) was used to count and extract the duplex reads in the resulting bam files. Barcode demultiplexing from duplex bam files was conducted with Dorado, resulting in one duplex bam file per sample. In this step, the barcodes were trimmed off the sequences. Using SAMtools, bam files were then transformed to fastq files. Using cutadapt v. 1.18 (Martin 2011), sequences < 1400 and > 1600 bp (Girija et al. 2023) were removed from the datasets. Due to noisy signals at the beginning and the end of each sequence, the first and the terminal 50 bases were removed from each sequence with cutadapt. Finally, we extracted all sequences with a mean Phred score ≥ 30 (Q3). As the dada2 pipeline used for ASV calling of Illumina reads is for several reasons not suitable for Nanopore's long-read 16S rRNA sequences (Zhang et al. 2023), sequences of the high-quality fastq files were binned into operational taxonomic units (OTUs) using NanoCLUST (Rodríguez-Pérez et al. 2021) as implemented in the BugSeq 16S pipeline (Fan et al. 2021; Jung and Chorlton 2021). The sequence similarity threshold for clustering sequences was 98%, corresponding to bacterial species demarcation (Kim et al. 2014).

Taxonomic assignments

Taxonomic classification of Illumina-derived ASVs and Nanopore-derived OTUs was conducted using the 16S BugSeq pipeline (Fan et al. 2021; Jung and Chorlton 2021). As the results of taxonomic classifications can be notably influenced by the methodology of assignment and by the reference database (Petrone et al. 2023; Zhang et al. 2023), we chose the same classification pipeline for both the Illumina ASVs and the Nanopore OTUs. BugSeq's 16S pipeline employs the QIIME2 VSEARCH consensus classifier (Rognes et al. 2016; Hall and Beiko 2018). Sequences were aligned against NCBI's nucleotide (nr) database with both sequence similarity and alignment length thresholds being set

at the default (80%), ensuring that only high-quality alignments are retained. In case of multiple alignments with an equal top score, equally good top hits were collapsed to their lowest common ancestor to ensure an as high as possible accuracy in taxonomic classification (Jung and Chorlton 2021).

Statistics

For pattern matching between the Illumina-derived ASV-to-sample matrix and the Nanopore-derived OTU-to-sample matrix, we calculated standard measures of diversity. Data analyses were conducted in R v. 4.0.5 using the packages vegan (Oksanen et al. 2022) and ggplot2 (Wickham 2016) for graphical visualisation. Prior to the calculation of alpha- and beta-diversity measures, the number of sequences per sample was rarefied (normalised) to the smallest sample size with the *rrarefy* function. In case of the Nanopore OTU matrix, this number was 10,986 sequences (LIS_REF, Table 1) and, in case of the Illumina ASV matrix, 60,607 (LIS_REF, Table 2). We calculated the Shannon-Wiener Index H' for both datasets as a measure of alpha-diversity. The abundance-based Bray-Curtis (BC) index and the incidence-based Jaccard index were used to calculate the similarity between samples (beta-diversity), based on the Hellinger transformed ASV and OTU data. For the calculation of both indices, we considered: (i) the complete ASV and OTU datasets; (ii) only ASVs and OTUs assigned to the taxonomic rank of genus; (iii) only ASVs and OTUs assigned to the taxonomic rank of family. In case of the complete ASV and OTU datasets, four separate distance matrices (ASV Jaccard, ASV BC, OTU Jaccard, OTU BC) had to be calculated, because ASVs and OTUs are taxonomy-independent and, therefore, the two datasets could not be superimposed. In case of beta-diversity indices based on taxonomic ranks, we united the ASV data and the OTU data to a single matrix, one that included all sequences assigned to the rank of families and one to the rank of genera. Thereby, we only considered families and genera in which the sequence read counts accounted for at least 0.5% of all sequence reads within an individual sample. For visualisation of beta-diversity, we then performed hierarchical clustering using the Ward criterion (Murtagh and Legendre 2014).

For a more in-depth taxonomic comparison, we compared the top ten (in terms of sequence reads) bacterial families and genera obtained from both sequence datasets.

Results

Overview of sequence data

The evolution of Nanopore sequence data from raw reads (pod5) per sample to high quality reads which could be taxonomically assigned to the domain Bacteria is shown in Table 1. From an initial nearly 5,200,000 reads after a runtime of 21 hours and 34 minutes (estimated bases: ca. 9 Gbp with an approximated N50 of 1.49 Kbp). The remaining total number of reads after a rigorous quality filtering (see Methods section above) for downstream analyses was 131,494 with the lowest number of reads being in sample LIS_REF (10,986) and the highest in sample DUN_AZE (41,580). The main loss of sequence reads occurred during the extraction of duplex reads and then again after barcode filtering: of the ca. 2 million

duplex reads, roughly 500,000 remained that could undoubtedly be assigned to the seven samples (six eDNA samples plus one non-template control sample). None of the samples had a singleton read. In the non-template control, we found eight sequences, none of which returned a hit after aligning to NCBIs nucleotide database. None of these eight sequences occurred in any of the final sample data.

In case of the Illumina dataset (Table 2), the loss of sequence reads after quality filtering was notably lower. Throughout the sequence data processing pipeline, we have lost a per-sample average of 43% of the originally obtained R1/R2 sequence reads. In numbers, we have started with roughly one million reads for the six eDNA samples of which we could use 448,036 for downstream analyses.

Table 1. This table shows the loss of Nanopore sequence reads from the original pod5 output to the per-sample high-quality reads with taxonomic classification to the domain Bacteria.

Sample	n total reads	n simplex reads	n duplex reads	Duplex reads per barcode	n reads passing length filter (1400-1600 bp)	n reads ≥ Phred30	n reads after taxonomic assignment**
LIS_CE	5,199,645*	3,243,488*	1,956,157*	27,654	20,566	12,954	11,512
LIS_AZE				103,534	81,262	32,188	28,234
LIS_REF				36,259	28,529	12,938	10,986
DUN_CE				77,133	61,412	26,329	23,687
DUN_AZE				141,389	104,983	45,583	41,580
DUN_REF				52,399	35,399	17,549	15,495
Control				280	16	8	0

* numbers refer to the complete dataset before sample-specific (barcode) demultiplexing.
** to the domain Bacteria.

Macrofaunal-based biotic index of LIS samples

The AZTI Marine Biotic Index (AMBI) obtained from the LIS samples in the framework of a compliance monitoring survey resulted in the following finding. Environmental quality was similarly high at LIS_REF (AMBI: 2.0 for replicate 1 and 1.8 for replicate 2) and LIS_AZE (AMBI: 2.3 for replicate 1 and 2.2 for replicate 2), while LIS-CE was notably more impacted (AMBI: 5.7 for replicate 1 and 5.8 for replicate 2).

Shannon diversity

The relative trend in Shannon diversity was largely congruent between the Illumina-derived ASV and the Nanopore-derived OTU datasets (Table 3). Both agreed that the lowest Shannon diversity for both the LIS and the DUN farm was at the cage edge sites (CE). Further commonalities between the two datasets were that the Shannon diversity at the CE and the reference (REF) sites were lower at the LIS farm compared to the DUN farm. Noteworthy differences in Shannon diversity between the two datasets were: first, the differences between the two aquaculture installations were notably more pronounced in the Nanopore dataset compared to the Illumina dataset. Second, both datasets disagreed in the obtained Shannon indices for the Allowable Zone of Effect (AZE). For both farms, the Illumina dataset suggested a Shannon diversity at the AZE which was in the same order of magnitude as the one obtained from the REF samples. In contrast, the Nanopore dataset revealed a notably higher Shannon diversity at the AZE compared to the REF for both farms.

Table 2. This table shows the loss of Illumina sequence reads from the original Illumina R1/R2 fastq output to the per-sample high-quality reads with taxonomic classification to the domain Bacteria.

Sample	R1/R2 output reads	Paired reads with Phred score ≥ 30	n reads with taxonomic classification*	n reads with taxonomic classification, singletons omitted**
LIS_CE	187,682	154,039	102,574	102,519
LIS_AZE	182,486	139,241	70,186	70,019
LIS_REF	155,450	127,725	60,712	60,607
DUN_CE	146,900	114,526	69,473	69,406
DUN_AZE	185,666	137,589	81,482	81,361
DUN_REF	180,007	133,086	64,324	64,124

* to the domain Bacteria.
** singletons refer to sequence reads with an abundance of one across all six samples.

Table 3. Shannon index calculated from the Nanopore OTU-to-sample matrix and the Illumina ASV-to-sample matrix.

Sample	Nanopore	Illumina
LIS_CE	3.71	4.16
LIS_AZE	4.24	4.94
LIS_REF	3.92	4.83
DUN_CE	4.13	4.33
DUN_AZE	4.36	4.84
DUN_REF	4.21	4.93

Beta-diversity

The Nanopore and the Illumina datasets showed full agreement when matching their abundance-based Bray-Curtis beta-diversity patterns on an OTU and ASV basis, respectively (Fig. 2). Both datasets showed two distinct clusters. One consisted of the cage edge samples of both aquaculture installation sites (LIS_CE and DUN_CE) plus the AZE sample from the DUN farm. In both cases, the DUN_CE and DUN_AZE were more similar to each other in their bacterial community profiles than any of these two to the bacterial community at the LIS_CE. The second cluster consisted of the reference samples of both farms (DUN_REF and LIS_REF) and the AZE sample of the LIS farm. In both cases, the LIS_REF and LIS_AZE were more similar to each other than any of the two was to the DUN_REF sample. One noteworthy difference was that the Bray-Curtis distances between the bacterial communities were higher in the Nanopore dataset (Fig. 2a) compared to the Illumina dataset (Fig. 2b).

When collating the OTU and ASV matrices on the taxonomic ranks family, genus and species, the pattern described above for ASVs and OTUs remained the same in case of the Nanopore dataset (Figs 2c, e, g). In case of the Illumina dataset, the beta-diversity patterns, based on the taxonomic ranks family (Fig. 2d) and genus (Fig. 2f), changed slightly in one aspect. While the two large clusters remained the same across all beta diversity analyses, the within-group clustering of the "DUN_CE – LIS_CE – LIS_AZE" cluster has changed in the Illumina dataset. Instead of a higher similarity between the bacterial communities of the LIS_REF and LIS_AZE as in all other beta diversity analyses, we
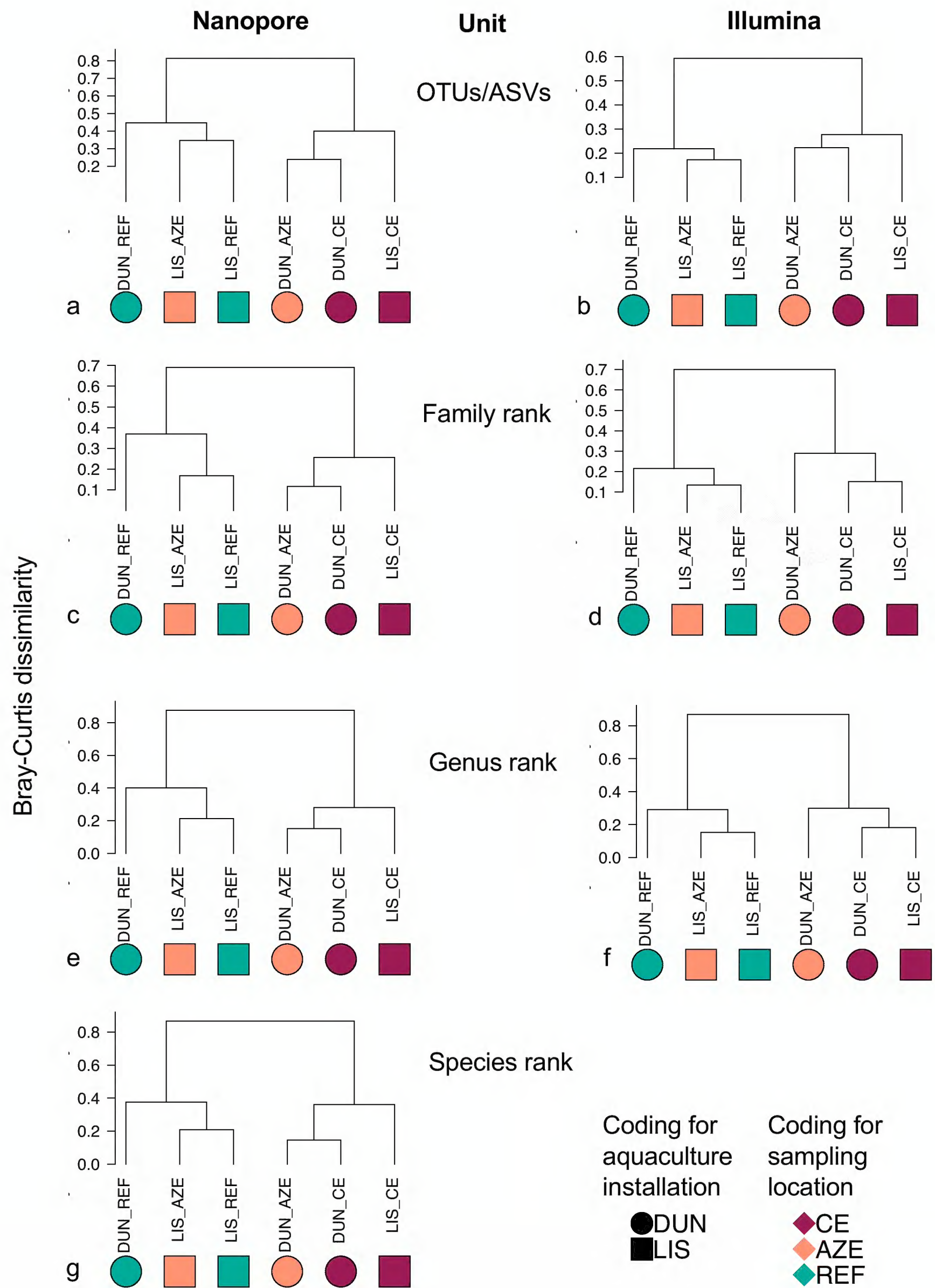
Figure 2. Hierarchical clustering (based on Bray Curtis index as a measure of beta diversity) as a measure of beta diversity, based on Nanopore-obtained amplicons (left panel) and Illumina-obtained amplicons (right panel) for the two salmon aquaculture installations DUN and LIS. The middle panel shows the (taxonomic) units on which the beta-diversity analyses are based. Colour and shape coding of samples helps visualisation and interpretation of data. AZE = Allowable Zone of Effect; CE = Cage Edge; REF = unimpacted Reference site.

found a higher similarity between the LIS_REF and DUN_REF samples in both taxonomy-based Illumina datasets.

While in case of the Nanopore dataset, sufficient bacterial species with high read abundances (> 0.5%) were identified (between 28 in samples LIS_CE and DUN_CE and 39 OTUs in sample DUN_REF, Table 4) to allow for a beta-diversity analysis, this was not the case for the Illumina dataset. Here, only between two (LIS_REF) and nine (LIS_CE) species could be identified. This was an insufficient number to conduct beta-diversity analyses on the species rank.

The beta-diversity results obtained for the incidence-based Jaccard index were identical to the ones described above for the abundance-based Bray-Curtis index (Suppl. material 1: figs S1–S7).

Taxonomic assignments

On average, 86% (± 2.8%) of the high-quality Nanopore amplicon reads could be assigned to the taxonomic rank of bacterial family, whereas this was only 52% (± 7.8%) of the high-quality Illumina dataset (Table 4). A maximum of only 13% (LIS_AZE and DUN_REF) of the total bacterial families inferred from Illumina amplicons had a sequence read abundance of > 0.5%. Thus, in the Illumina dataset, the vast majority of families obtained had low read abundances (Table 4). In case of the Nanopore dataset, up to 84% (LIS_CE) of the identified families had a read abundance of > 0.5%. An average of 82% (± 1.8%) of the Nanopore reads could be assigned to the taxonomic rank of bacterial genus, which was only 50% (± 7.1%) in case of the Illumina sequence reads (Table 4). With very few exceptions, all the Illumina-derived genera had low read abundances. In case of the Nanopore dataset, on average, more than half (54% ± 11%) of the detected genera were with high-abundant reads (Table 4). In the Nanopore datasets, a per-sample average of 65% (± 11%) of all sequence reads returned a species hit in the taxonomic analysis, whereas this was only for an average of 15% (± 4%) of the Illumina sequence reads. In the Nanopore dataset, up to 72% of the identified bacterial species had high-abundant reads, whereas this applied to a maximum of only 3% of the Illumina-derived species. Up to > 99% of the ASVs obtained from Illumina sequencing that were assigned a species rank belonged to the rare ASVs.

Seven out of the ten most abundant Nanopore-derived families (combined over all six samples) were also amongst the ten most abundant bacterial families of the Illumina dataset, albeit with distinct relative abundances (Fig. 3). These shared families were Desulfosarcinaceae, Desulfocapsaceae, Desulfobulbaceae, Desulfobacteraceae, Sulfurovaceae, Woeseiaceae and Halieaceae. Thiotrichaceae, Sedimenticolaceae and Sandarinaceae were amongst the top ten Nanopore-derived families, but not amongst the top ten Illumina-derived families.

While Sedimenticolaceae and Sandarinaceae were also present in the Illumina dataset, albeit not amongst the ten most abundant families, Thiotrichaceae entirely escaped Illumina detection (Fig. 4). Chromatiaceae, Thioprofundaceae and Prolixibacteraceae were amongst the top ten Illumina-derived families. While Chromatiaceae was also recorded with the Nanopore protocol, albeit not amongst the ten most abundant, the latter two were not recorded at all with Nanopore sequencing (Fig. 4). In general, only few families which accounted for > 0.5% of the sequence reads in one of the datasets was entirely missing from the other. In addition to Thioprofundaceae and

Table 4. Proportions of Nanopore and Illumina sequence reads that could be assigned to the taxonomic ranks family, genus and species. Proportional numbers refer to the proportion of all reads of an individual sample that could be assigned to each of these taxonomic ranks regardless of the sequence read abundance assigned to an individual taxon and to taxa which had a read abundance of at least 0.5% (= "high abundant" taxa).

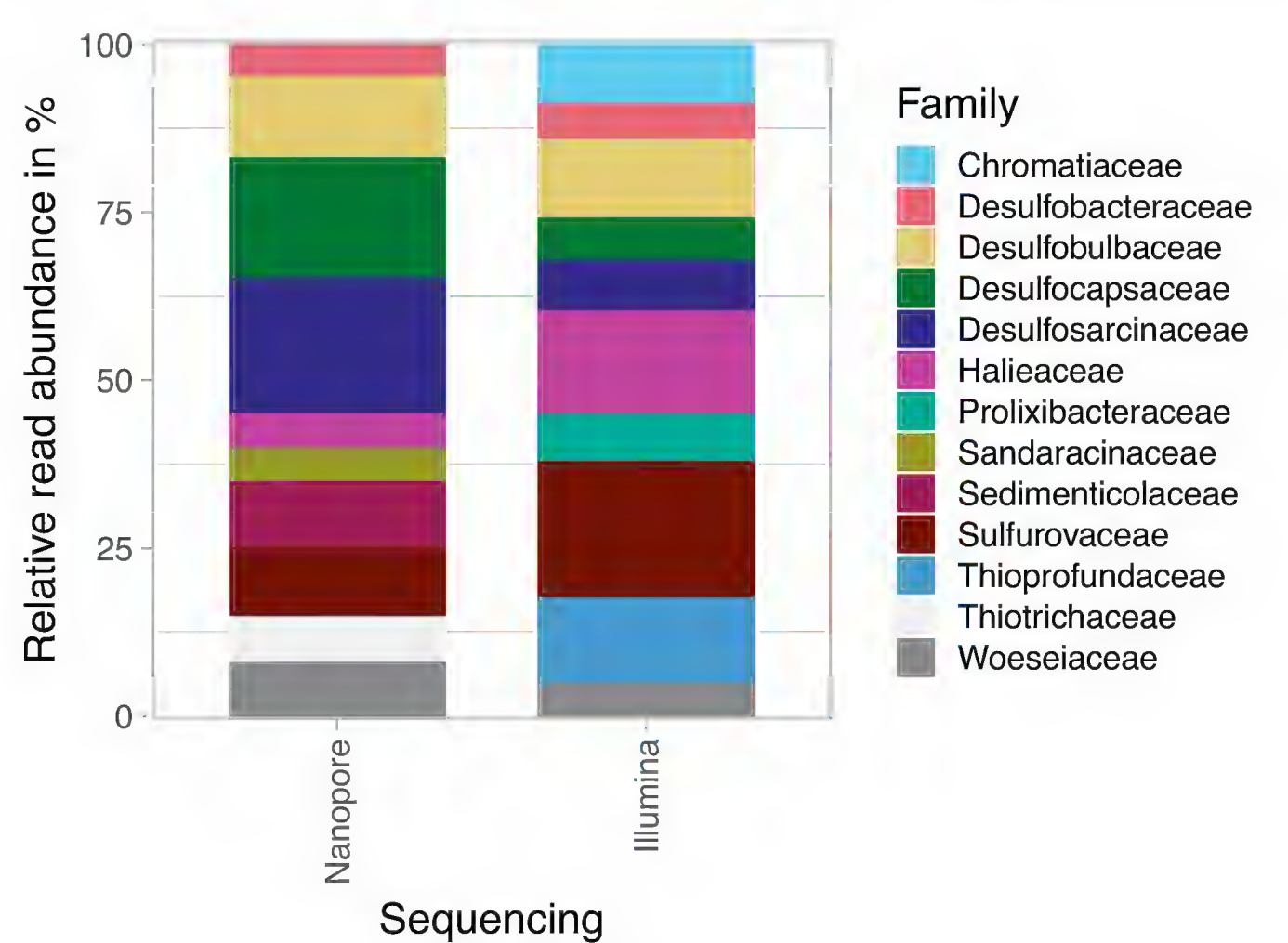| Sequencing platform	Sample	% reads assigned to rank family	% of families which have a read abundance of > 0.5%	% reads assigned to rank genus	% of genera which have a read abundance of > 0.5%	% reads assigned to rank species	% of species which have a read abundance of > 0.5%
Nanopore	LIS_CE	88	84	81	69	59	72
Nanopore	LIS_AZE	86	54	82	49	71	46
Nanopore	LIS_REF	83	75	83	65	68	76
Nanopore	DUN_CE	89	65	84	55	59	49
Nanopore	DUN_AZE	90	48	85	46	62	34
Nanopore	DUN_REF	83	49	80	43	72	66
Illumina	LIS_CE	63	9	61	6	19	3
Illumina	LIS_AZE	47	13	45	5	12	1
Illumina	LIS_REF	44	12	42	5	11	0
Illumina	DUN_CE	58	11	56	6	18	2
Illumina	DUN_AZE	56	11	53	5	18	2
Illumina	DUN_REF	45	13	43	4	12	1

Figure 3. The ten most abundant (in terms of sequence reads) bacterial families and their relative abundances obtained from the Nanopore and Illumina datasets.

Prolixibacteraceae, two further bacterial families, Desulfosalsimonadaceae and Wenzhouxiangellaceae, were exclusively present in the Illumina dataset, but missing from the Nanopore dataset. In addition to Thiotrichaceae, Thermoanerobaculaceae and Spirochaetaceae escaped detection with the Illumina protocol, but were detected with Nanopore sequencing (Fig. 4).

Comparing the ten genera with the most abundant reads in both sequence datasets (Fig. 5), we found four genera that were shared. These were *Desulfobulbus*, *Sulfurovum*, *Halioglobus* and *Woesia*. The genera *Desulfatiglans*,
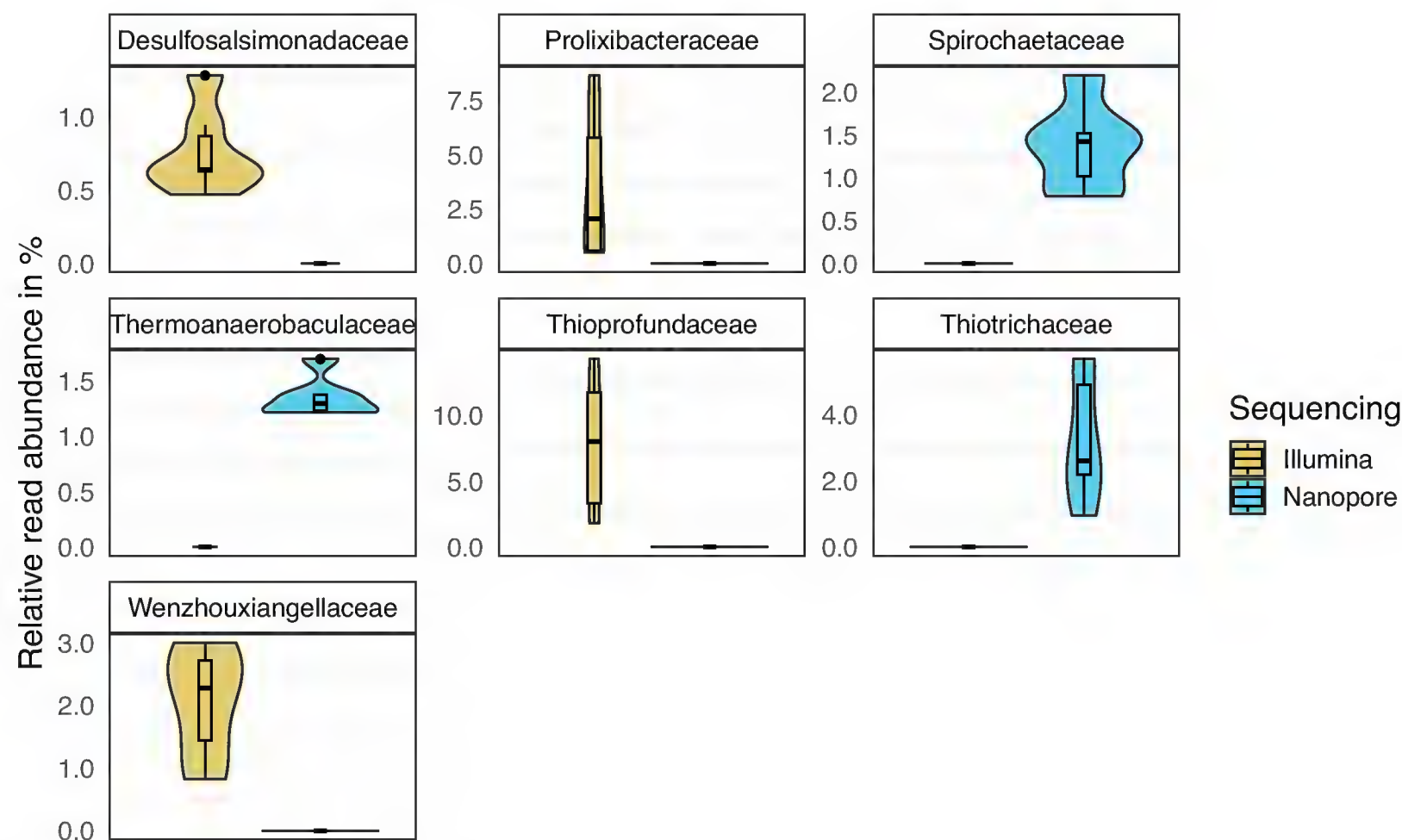
Figure 4. Bacterial families which were detected exclusively with Nanopore or with Illumina sequencing. The violin-and-box-plots show the relative abundances. Violins (coloured areas) show the relative abundance distribution across the individual samples. Boxes show median, 25%- and 75%-quartiles and min-max values.
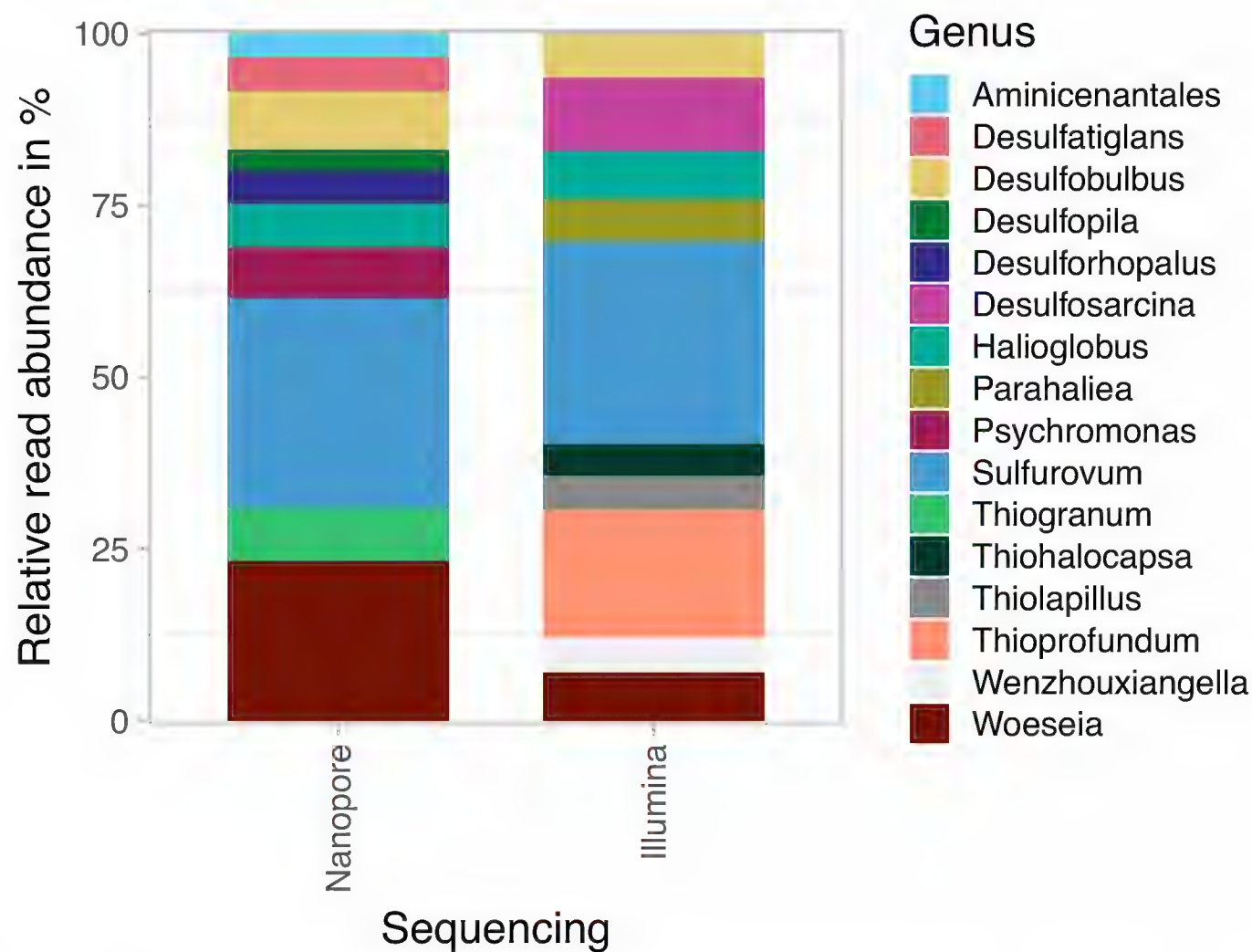

Figure 5. The ten most abundant (in terms of sequence reads) bacterial genera and their relative abundances obtained from the Nanopore and Illumina datasets.

Desulforhopalus, *Desulfopila*, *Thiogranum*, *Psychromonas* and *Aminicenantales* were exclusively amongst the top ten of the Nanopore dataset (Fig. 5). With the exceptions of *Thiogranum* and *Aminicenantales*, these genera were also recorded with Illumina sequencing; however, not amongst the ten most abundant

genera. *Desulfosarcina*, *Parahaliea*, *Thiohalocapsa*, *Thiolapillus*, *Thioprofundum* and *Wenzhouxiangella* were amongst the ten most abundant genera in the Illumina dataset, but not amongst the top ten of the Nanopore dataset.

With the exception of *Desulfosarcina*, none of the other five Illumina-derived top ten genera was recorded with the Nanopore dataset. Only one genus (*Thiogranum*) which had a sequence read abundance of > 0.5% in the Nanopore dataset escaped detection with the Illumina dataset (Fig. 6). In contrast, five genera (*Wenzhouxiangella*, *Thioprofundum*, *Thiolapillus*, *Thiohalocapsa*, *Parahaliea* and *Marimicrobium*) with a sequence read abundance of > 0.5% in the Illumina dataset could not be detected with the Nanopore protocol (Fig. 6)
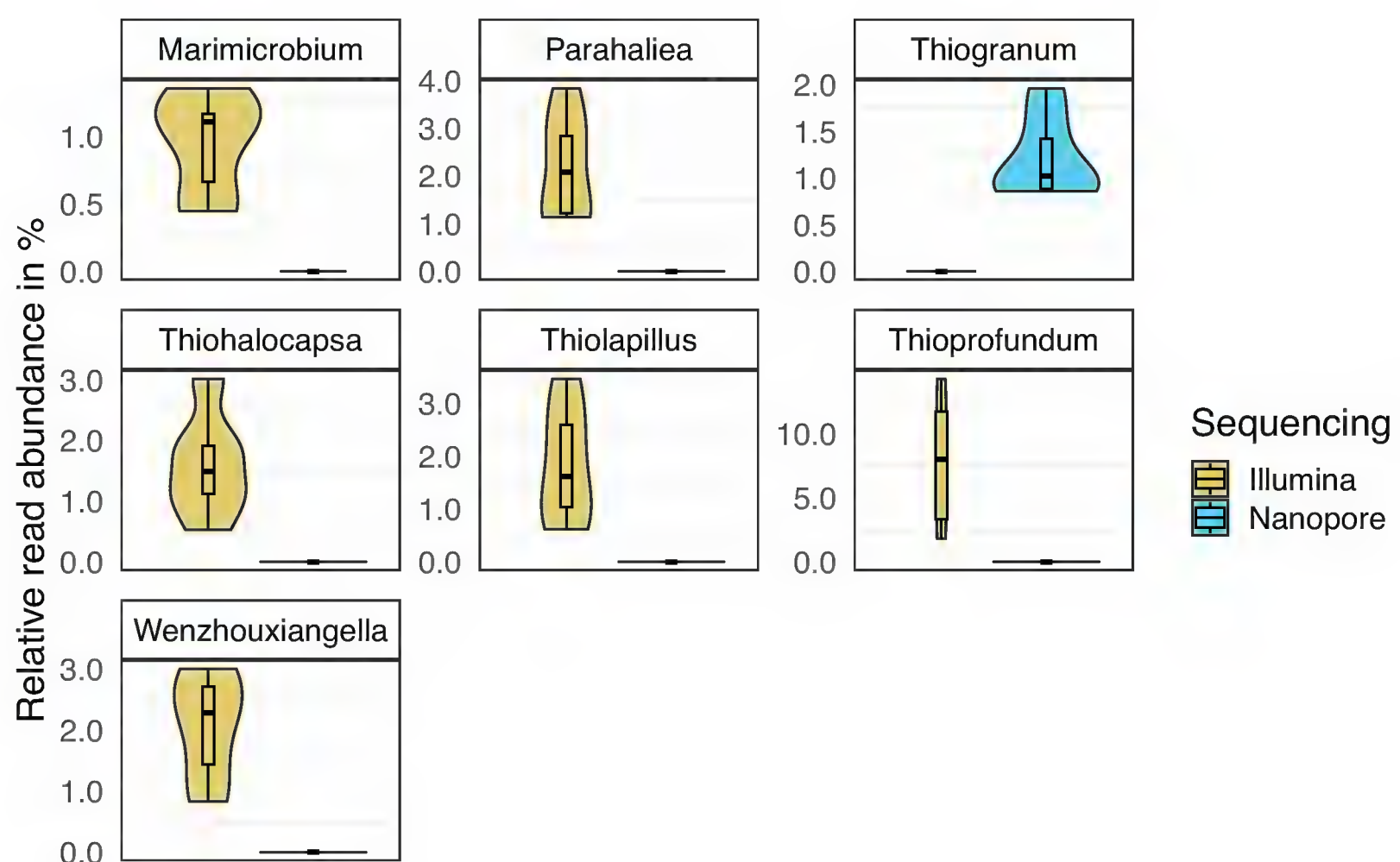

Figure 6. Bacterial genera which were detected exclusively with Nanopore or with Illumina sequencing. The violin-and-box-plots show the relative abundances. Violins (coloured areas) show the relative abundance distribution across the individual samples. Boxes show median, 25%- and 75%-quartiles and min-max values.

Discussion

Data quality of Nanopore duplex reads

Our results witness a notably improved data quality of Nanopore sequences compared to previous studies (e.g. Older et al. (2023); Petrone et al. (2023); Zhang et al. (2023)). This improved sequence quality is a result of the recently introduced sequencing chemistry, which, in combination with the new R10.4.1 flow cell, enables duplex reads, whereas the previous chemistry with the same flow cell allowed only for single-molecule reads (simplex reads) only. For example, in our LIS_CE dataset, 63% of the duplex reads with the correct length have passed the Phred30 quality filter. Previous Nanopore sequencing studies analysing bacterial mock or natural communities did not even attempt to use such a high Phred score quality filter as is standard for Illumina MiSeq data. Older et al. (2023) found an average of 59% of 16S rRNA gene sequence reads that have

passed a Phred20 filter. Petrone et al. (2023) reported an average Phred score of 18.1 with an average of 79% of reads passing a Phred15 filter. Similarly, Zhang et al. (2023) obtained an average Phred score of 18.8 for their bacterial 16S rRNA gene dataset, but did not apply any Phred quality filter for downstream analyses.

The proportion of duplex reads obtained in our study was 36% (Table 1) and, thus, notably higher compared to other studies. Genome-sequencing different bacterial species, Lerminiaux et al. (2024), as well as Sanderson et al. (2023), reported a duplex rate of only ca. 6% and 7%, respectively, using the same chemistry and flow cell type we have used in our study. It is not unlikely that the success rate of duplex sequencing correlates with the length of the target gene. Nanopore genome sequencing produces up to > 2 Mbp, with 10–30 Kbp being common (Amarasinghe et al. 2020; Lerminiaux et al. 2024). There may be a higher likelihood for shorter DNA fragments binding at the pore while waiting to follow the first strand through the pore. This assumption, however, needs to be verified. Thus far, no further study is published (to the best of our knowledge) that has analysed bacterial community structures using 16S rRNA gene Nanopore duplex sequencing (or any other targeted gene with a size < 2Kbp), which would allow a confirmation of this assumption.

In addition, while Sanderson et al. (2023), as well as Lerminiaux et al. (2024), have used ONTs Guppy tool for duplex base-calling, we in this study have used ONTs new Dorado base-caller, which relies on a bi-directional Recurrent Neural Network (RNN) algorithm that was optimised for duplex reads in contrast to the Guppy base-caller (https://github.com/nanoporetech/dorado, Wick (2023)). Despite a high loss of sequence reads during duplex base-calling and another noteworthy loss due to erroneous barcode reads, the data per sample that we have retained for analyses exceeded by far the number of reads required for a sample when subjected to DNA-based marine biomonitoring. Dully et al. (2021b) tested several different marine DNA-based biomonitoring scenarios using microbial communities and identified a per-sample sequence depth of 3,000–5,000 sequences as sufficient. Any further increase in sequence numbers did not affect the monitoring results. Wilding et al. (2023) analysed the signal-to-noise ratio in 16S rRNA gene amplicon dataset obtained from DNA-based monitoring of coastal salmon aquaculture installations. The authors found that between 10 and 100 top abundant bacterial ASVs are optimal for biomonitoring purposes. Both recommendations are met in our Nanopore sequence datasets, thus, allowing for a comparison of the results obtained by Nanopore sequencing with the ones obtained from the sequencing of the same samples with Illumina MiSeq.

Comparison of ecological trends inferred from Nanopore- and Illumina-derived bacterial diversity patterns

Both sequencing protocols produced largely consistent alpha- and beta-diversity results. This agrees with other studies which also observed consistent ecological trends in 16S rRNA gene data obtained from the two sequencing platforms (e.g. Nygaard et al. (2020); Lemoinne et al. (2023); Older et al. (2023); Stevens et al. (2023); Zorz et al. (2023)). On the level of OTUs (Nanopore) and ASVs (Illumina), both datasets showed the same clustering of the six different samples collected from the two aquaculture

installations. The two main clusters grouped the samples according to environmental impact into a "high impact cluster" (samples DUN_CE, DUN_AZE, LIS_CE) and a "low impact cluster" (samples DUN_REF, LIS_REF, LIS_AZE). This corroborates well with previous findings which suggested that the structures of benthic bacterial communities are a robust indicator of environmental impact arising from fish farming (Dowle et al. 2015; Keeley et al. 2018; Stoeck et al. 2018; Dully et al. 2021a; Frühe et al. 2021b; Leontidou et al. 2023; Wilding et al. 2023).

Both LIS_REF and LIS_AZE had AMBI values, obtained from macroinvertebrate-based compliance monitoring of the LIS farm, which indicated a good ecological status (Muxika et al. 2005). Thus, we can conclude that the LIS_AZE was hardly impacted by organic pollution through the salmon farm. This explains the clustering of LIS_AZE together with the two reference samples of both farms (LIS_REF and DUN_REF). These reference samples were chosen at a distance far enough from the fish pens to remain unimpacted by any deposits resulting from the aquaculture activities (Wilding et al. 2023). Furthermore, we can conclude that the DUN_AZE was notably more impacted than the LIS_AZE, because DUN_AZE together with DUN_CE clustered together with LIS_CE, which had an averaged AMBI of 5.75, indicating a bad ecological status and a heavily disturbed site (Muxika et al. 2005). A further interesting finding from the OTU/ASV-based beta-diversity analyses is that the Bray-Curtis distances between the bacterial communities were higher in the Nanopore datasets compared to the Illumina dataset. This suggests a higher sensitivity of the Nanopore dataset in distinguishing bacterial communities of different samples. This finds support in the results obtained for the Shannon index. Even though the general trend is highly similar, the differences between individual samples are notably more pronounced in the Nanopore data compared to the Illumina data. A higher sensitivity in bacterial community diagnostics is of advantage in biomonitoring to indicate more subtle environmental changes.

Low taxonomic resolution of short Illumina reads affects taxonomy-based bacterial diversity patterns

When collating sequence data across the taxonomic levels family and genus, the branching of samples within the Illumina-derived high-impact clusters changed compared to all other beta-diversity analyses. Short hypervariable 16S rRNA gene regions often lack a differentiation between genera of the same family and, likewise, between species of the same genus (Callahan et al. 2019; Klair et al. 2023; Older et al. 2023; Zhang et al. 2023). Therefore, it is not surprising that several studies reported differences in the presence/absence of individual taxa in bacterial communities that were sequenced with both Illumina and Nanopore (Nygaard et al. 2020; Klair et al. 2023; Lemoinne et al. 2023; Stevens et al. 2023; Tandon et al. 2023). In the Nanopore dataset, we only considered sequences with ≥ 98% similarity to a deposited sequence in the NCBI database, which allows for more accurate taxonomic classifications compared to short reads (Benitez-Paez et al. 2016; Klair et al. 2023; Zhang et al. 2023). More than one third (36%) of the here-obtained high-quality Nanopore sequences used in downstream analyses had a sequence similarity of > 99% to NCBI reference sequences which corresponds to the species- and strain-discrimination

boundaries in bacterial 16S rRNA gene sequences (Benitez-Paez et al. 2016). This makes taxonomic classifications, based on long Nanopore reads, more robust and reliable compared to the short Illumina reads, particularly on lower taxonomic levels.

The low taxonomic resolution of short Illumina reads is furthermore of disadvantage regarding the effect of sequencing errors on taxonomic assignment accuracy. A MiSeq-platform inherent weakness is the accumulation of substitution errors (Schirmer et al. 2015; Schirmer et al. 2016). This may lead to sequences that do not exist in nature and, eventually to spurious taxa or ASVs (false positives) which logically cannot be detected with Nanopore sequencing (Xue et al. 2018; Stevens et al. 2023). Additionally, Nanopore sequencing is generating errors. However, if the same Phred sequence quality filter is applied, the effects of these errors on taxonomic assignment accuracy are much lower for the (near) full-length 16S rRNA gene compared to a short gene fragment.

PCR primer bias and stochasticity may affect analyses of bacterial community structures obtained from Nanopore and Illumina sequencing

Different protocols for sequence generation may distort relative abundance comparisons, in particular due to PCR primer bias and stochasticity (Balint et al. 2016; Lemoinne et al. 2023). The primer pair we used in this study to amplify the V3-V4 16S rRNA gene region is very popular in bacterial diversity research in combination with Illumina MiSeq sequencing. Amongst others, this primer pair is frequently used as standard in biomonitoring of aquaculture installations (Stoeck et al. 2018; Dully et al. 2021a; Dully et al. 2021b; Frühe et al. 2021a; Frühe et al. 2021b; Wilding et al. 2023). A recent study demonstrated that numerous bacterial taxon groups, amongst others on family and genus level, but even on phylum-level, are missed by this primer pair (Leontidou et al. 2023). For the amplification of the full-length 16S rRNA gene for Nanopore sequencing, we also used a primer pair (27F/1492R) which is very popular in bacterial diversity research (e.g. Galkiewicz and Kellogg (2008); Klindworth et al. (2013); Johnson et al. (2019); Fujiyoshi et al. (2020); Older et al. (2023); Tandon et al. (2023)). Likewise, this primer pair 27F/1492R, or variants thereof, discriminates against several bacterial taxon groups as demonstrated in previous studies (see, for example, Older et al. (2023); Tandon et al. (2023)).

In addition, differences in relative read abundances typically affect ecological trends inferred from Nanopore and Illumina sequencing of the same bacterial (mock and natural) communities (e.g. Klair et al. (2023); Older et al. (2023); Stevens et al. (2023); Tandon et al. (2023)). Many diversity measures, used with ASVs or OTUs, such as the Bray Curtis Index, are abundance-weighted and, therefore, largely influenced by the most abundant ASVs or OTUs rather than by the less abundant ones (Chiarello et al. 2022). Differences in relative abundances of individual bacterial taxon groups reflect an inherent technical bias of all PCR-based high-throughput sequencing methods resulting from the stochastic process of target-gene amplification (Balint et al. 2016; Lemoinne et al. 2023). One possibility to exclude the two above-mentioned sources of error that apply equally to Nanopore and Illumina amplicon protocols is PCR-independent metagenome sequencing (Leontidou et al. 2023). This is, however, too expensive to apply in routine DNA-based biomonitoring practice.

Conclusions

Nanopore amplicon duplex sequencing is a very promising alternative to Illumina sequencing to analyse bacterial community structures in sediments collected from salmon aquaculture installations. Ecological trends inferred from Nanopore- and Illumina-derived bacterial OTUs/ASVs are highly similar. Taxonomy-based ecological analyses of Nanopore 16S rRNA gene data are more reliable and trustworthy than the ones obtained with short Illumina reads of hypervariable 16S rRNA gene regions. This is in particular due to: (i) the higher robustness of the long Nanopore reads to errors compared to the short Illumina reads when the same Phred quality filter is applied and (ii) the high taxonomic resolution power of the (near) complete 16S rRNA gene. The obtained results of this pilot study provide confidence to move to the next step: the assessment of Nanopore amplicon duplex sequencing the performance, robustness and accuracy in compliance monitoring of aquaculture installations. As accomplished previously to assess the performance of Illumina sequencing in compliance monitoring (Keeley et al. 2018; Cordier et al. 2019; Dully et al. 2021a; Frühe et al. 2021b; Wilding et al. 2023), this will require the analyses of numerous samples with a known biotic index inferred from compliance monitoring protocols in place.

The R scripts used in this study were deposited in GitHub (https://github.com/verubel/CoastMon/tree/main/Nanopore_vs_Illumina).

Acknowledgements

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Funding

Author contributions

Conceptualization: TS. Data curation: TS. Formal analysis: TS, SNK, HWB, VR. Funding acquisition: TS. Investigation: TS. Methodology: TS. Project administration: TS. Validation: VR, TS, SNK. Visualization: SNK. Writing - original draft: TS. Writing - review and editing: HWB, VR, SNK.

Author ORCIDs

Thorsten Stoeck ⓘ https://orcid.org/0000-0001-5180-5659
Verena Rubel ⓘ https://orcid.org/0000-0001-9630-9050

Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

References

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q (2020) Opportunities and challenges in long-read sequencing data analysis. Genome Biology 21(1): 30. https://doi.org/10.1186/s13059-020-1935-5

Aylagas E, Borja A, Tangherlini M, Dell'Anno A, Corinaldesi C, Michell CT, Irigoien X, Danovaro R, Rodriguez-Ezpeleta N (2017) A bacterial community-based index to assess the ecological status of estuarine and coastal environments. Marine Pollution Bulletin 114(2): 679–688. https://doi.org/10.1016/j.marpolbul.2016.10.050

Balint M, Bahram M, Eren AM, Faust K, Fuhrman JA, Lindahl B, O'Hara RB, Opik M, Sogin ML, Unterseher M, Tedersoo L (2016) Millions of reads, thousands of taxa: Microbial community structure and associations analyzed via marker genes. FEMS Microbiology Reviews 40(5): 686–700. https://doi.org/10.1093/femsre/fuw017

Bannister RJ, Valdemarsen T, Hansen PK, Holmer M, Ervik A (2014) Changes in benthic sediment conditions under an Atlantic salmon farm at a deep, well-flushed coastal site. Aquaculture Environment Interactions 5(1): 29–47. https://doi.org/10.3354/aei00092

Benitez-Paez A, Portune KJ, Sanz Y (2016) Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION portable nanopore sequencer. GigaScience 5(1): 4. https://doi.org/10.1186/s13742-016-0111-z

Borja A (2018) Testing the efficiency of a bacterial community-based index (microgAMBI) to assess distinct impact sources in six locations around the world. Ecological Indicators 85: 594–602. https://doi.org/10.1016/j.ecolind.2017.11.018

Borja A, Franco J, Perez V (2000) A marine Biotic Index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Marine Pollution Bulletin 40(12): 1100–1114. https://doi.org/10.1016/S0025-326X(00)00061-8

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7): 581–583. https://doi.org/10.1038/nmeth.3869

Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, McGill SK, Dougherty MK (2019) High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Research 47(18): e103. https://doi.org/10.1093/nar/gkz569

Chen Z, Jiang Y, Dunphy DR, Adams DP, Hodges C, Liu N, Zhang N, Xomeritakis G, Jin X, Aluru NR, Gaik SJ, Hillhouse HW, Brinker JC (2010) DNA translocation through an array of kinked nanopores. Nature Materials 9(8): 667–675. https://doi.org/10.1038/nmat2805

Chiarello M, McCauley M, Villeger S, Jackson CR (2022) Ranking the biases: The choice of OTUs vs. ASVs in 16S rRNA amplicon data analysis has stronger effects on diversity measures than rarefaction and OTU identity threshold. PLoS ONE 17(2): e0264443. https://doi.org/10.1371/journal.pone.0264443

Cordier T, Lanzen A, Apotheloz-Perret-Gentil L, Stoeck T, Pawlowski J (2019) Embracing Environmental Genomics and Machine Learning for Routine Biomonitoring. Trends in Microbiology 27(5): 387–397. https://doi.org/10.1016/j.tim.2018.10.012

Cordier T, Alonso-Saez L, Apotheloz-Perret-Gentil L, Aylagas E, Bohan DA, Bouchez A, Chariton A, Creer S, Frühe L, Keck F, Keeley N, Laroche O, Leese F, Pochon X, Stoeck T, Pawlowski J, Lanzen A (2021) Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. Molecular Ecology 30(13): 2937–2958. https://doi.org/10.1111/mec.15472

Dawson SC, Pace NR (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. Proceedings of the National Academy of Sciences of the United States of America 99(12): 8324–8329. https://doi.org/10.1073/pnas.062169599

Delahaye C, Nicolas J (2021) Sequencing DNA with nanopores: Troubles and biases. PLoS ONE 16(10): e0257521. https://doi.org/10.1371/journal.pone.0257521

Dowle E, Pochon X, Keeley N, Wood SA (2015) Assessing the effects of salmon farming seabed enrichment using bacterial community diversity and high-throughput sequencing. FEMS Microbiology Ecology 91(8): fiv089. https://doi.org/10.1093/femsec/fiv089

Dully V, Rech G, Wilding TA, Lanzen A, MacKichan K, Berrill I, Stoeck T (2021a) Comparing sediment preservation methods for genomic biomonitoring of coastal marine ecosystems. Marine Pollution Bulletin, 173(Pt B): 113129. https://doi.org/10.1016/j.marpolbul.2021.113129

Dully V, Wilding TA, Mühlhaus T, Stoeck T (2021b) Identifying the minimum amplicon sequence depth to adequately predict classes in eDNA-based marine biomonitoring using supervised machine learning. Computational and Structural Biotechnology Journal 19: 2256–2268. https://doi.org/10.1016/j.csbj.2021.04.005

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Research 8(3): 186–194. https://doi.org/10.1101/gr.8.3.186

Fan J, Huang S, Chorlton SD (2021) BugSeq: A highly accurate cloud platform for long-read metagenomic analyses. BMC Bioinformatics 22(1): 160. https://doi.org/10.1186/s12859-021-04089-5

Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ (2008) Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. Applied and Environmental Microbiology 74(8): 2461–2470. https://doi.org/10.1128/AEM.02272-07

Frühe L, Cordier T, Dully V, Breiner HW, Lentendu G, Pawlowski J, Martins C, Wilding TA, Stoeck T (2021a) Supervised machine learning is superior to indicator value inference in monitoring the environmental impacts of salmon aquaculture using eDNA metabarcodes. Molecular Ecology 30(13): 2988–3006. https://doi.org/10.1111/mec.15434

Frühe L, Dully V, Forster D, Keeley NB, Laroche O, Pochon X, Robinson S, Wilding TA, Stoeck T (2021b) Global trends of benthic bacterial diversity and community composition along organic enrichment gradients of salmon farms. Frontiers in Microbiology 12: 637811. https://doi.org/10.3389/fmicb.2021.637811

Fujiyoshi S, Muto-Fujita A, Maruyama F (2020) Evaluation of PCR conditions for characterizing bacterial communities with full-length 16S rRNA genes using a portable nanopore sequencer. Scientific Reports 10(1): 12580. https://doi.org/10.1038/s41598-020-69450-9

Galkiewicz JP, Kellogg CA (2008) Cross-kingdom amplification using bacteria-specific primers: Complications for studies of coral microbial ecology. Applied and Environmental Microbiology 74(24): 7828–7831. https://doi.org/10.1128/AEM.01303-08

Girija GK, Tseng L-C, Chen Y-L, Meng P-J, Hwang J-S, Ho Y-N (2023) Microbiome variability in invasive coral (Tubastraea aurea) in response to diverse environmental stressors. Frontiers in Marine Science 10: 1234137. https://doi.org/10.3389/fmars.2023.1234137

Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: Ten years of next-generation sequencing technologies. Nature Reviews. Genetics 17(6): 333–351. https://doi.org/10.1038/nrg.2016.49

Hall M, Beiko RG (2018) 16S rRNA Gene Analysis with QIIME2. Methods in Molecular Biology (Clifton, N.J.) 1849: 113–129. https://doi.org/10.1007/978-1-4939-8728-3_8

Herlemann DP, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. The ISME Journal 5(10): 1571–1579. https://doi.org/10.1038/ismej.2011.41

Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, Leopold SR, Hanson BM, Agresta HO, Gerstein M, Sodergren E, Weinstock GM (2019) Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nature Communications 10(1): 5029. https://doi.org/10.1038/s41467-019-13036-1

Jung A, Chorlton SD (2021) BugSeq 16S: NanoCLUST with improved consensus sequence classification. bioRxiv 2021.03.16.434153. https://doi.org/10.1101/2021.03.16.434153

Karst SM, Ziels RM, Kirkegaard RH, Sørensen EA, McDonald D, Zhu Q, Knight R, Albertsen M (2021) High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nature Methods 18(2): 165–169. https://doi.org/10.1038/s41592-020-01041-y

Keeley NB, Forrest BM, Crawford C, Macleod CK (2012) Exploiting salmon farm benthic enrichment gradients to evaluate the regional performance of biotic indices and environmental indicators. Ecological Indicators 23: 453–466. https://doi.org/10.1016/j.ecolind.2012.04.028

Keeley N, Wood SA, Pochon X (2018) Development and preliminary validation of a multi-trophic metabarcoding biotic index for monitoring benthic organic enrichment. Ecological Indicators 85: 1044–1057. https://doi.org/10.1016/j.ecolind.2017.11.014

Kim M, Oh H-S, Park S-C, Chun J (2014) Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. International Journal of Systematic and Evolutionary Microbiology 64(Pt 2): 346–351. https://doi.org/10.1099/ijs.0.059774-0

Klair D, Dobhal S, Ahmad A, Hassan ZU, Uyeda J, Silva J, Wang KH, Kim S, Alvarez AM, Arif M (2023) Exploring taxonomic and functional microbiome of Hawaiian stream and spring irrigation water systems using Illumina and Oxford Nanopore sequencing platforms. Frontiers in Microbiology 14: 1039292. https://doi.org/10.3389/fmicb.2023.1039292

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research 41(1): e1. https://doi.org/10.1093/nar/gks808

Lane D (1991) 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (Eds) Nucleic acid techniques in bacterial systematics. John Wiley and Sons, Chichester-Toronto-New York-Brisbane-Singapore, 115–176.

Lanzén A, Mendibil I, Borja A, Alonso-Saez L (2021) A microbial mandala for environmental monitoring: Predicting multiple impacts on estuarine prokaryote communities of the Bay of Biscay. Molecular Ecology 30(13): 2969–2987. https://doi.org/10.1111/mec.15489

Laroche O, Wood SA, Tremblay LA, Ellis JI, Lear G, Pochon X (2018) A cross-taxa study using environmental DNA/RNA metabarcoding to measure biological impacts of offshore oil and gas drilling and production operations. Marine Pollution Bulletin 127: 97–107. https://doi.org/10.1016/j.marpolbul.2017.11.042

Lemoinne A, Guillaume D, Myriam G, Tony R (2023) Fine-scale congruence in bacterial community structure from marine sediments sequenced by short-reads on Illumina and long-reads on Nanopore. bioRxiv 2023.2006.2006.541006. https://doi.org/10.1101/2023.06.06.541006

Leontidou K, Abad-Recio IL, Rubel V, Filker S, Däumer M, Thielen A, Lanzén A, Stoeck T (2023) Simultaneous analysis of seven 16S rRNA hypervariable gene regions increases efficiency in marine bacterial diversity detection. Environmental Microbiology 25(12): 3484–3501. https://doi.org/10.1111/1462-2920.16530

Lerminiaux NA, Fakharuddin K, Mulvey M, Mataseje L (2024) Do we still need Illumina sequencing data?: Evaluating Oxford Nanopore Technologies R10.4.1 flow cells and the Rapid v14 library prep kit for Gram negative bacteria whole genome assemblies. Canadian Journal of Microbiology Epub ahead of print. https://doi.org/10.1139/cjm-2023-0175

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England) 25(16): 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

MacKenzie M, Argyropoulos C (2023) An introduction to Nanopore sequencing: Past, present, and future considerations. Micromachines 14(2): 459. https://doi.org/10.3390/mi14020459

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal 17(1): 10–12. https://doi.org/10.14806/ej.17.1.200

Murtagh F, Legendre P (2014) Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? Journal of Classification 31(3): 274–295. https://doi.org/10.1007/s00357-014-9161-z

Muxika I, Borja A, Bonne W (2005) The suitability of the marine biotic index (AMBI) to new impact sources along European coasts. Ecological Indicators 5(1): 19–31. https://doi.org/10.1016/j.ecolind.2004.08.004

Netzer R, Ribicic D, Aas M, Cave L, Dhawan T (2021) Absolute quantification of priority bacteria in aquaculture using digital PCR. Journal of Microbiological Methods 183: 106171. https://doi.org/10.1016/j.mimet.2021.106171

Nygaard AB, Tunsjo HS, Meisal R, Charnock C (2020) A preliminary study on the potential of Nanopore MinION and Illumina MiSeq 16S rRNA gene sequencing to characterize building-dust microbiomes. Scientific Reports 10(1): 3209. https://doi.org/10.1038/s41598-020-59771-0

Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara RB, Solymos P, Stevens MHH, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Weedon J (2022) vegan: Community Ecology Package. R package version 2.6-4. https://CRAN.R-project.org/package=vegan

Oladi M, Leontidou K, Stoeck T, Shokri MR (2022) Environmental DNA-based profiling of benthic bacterial and eukaryote communities along a crude oil spill gradient in a coral reef in the Persian Gulf. Marine Pollution Bulletin 184: 114143. https://doi.org/10.1016/j.marpolbul.2022.114143

Older CE, Yamamoto FY, Griffin MJ, Ware C, Heckman TI, Soto E, Bosworth BG, Waldbieser GC (2023) Comparison of high-throughput sequencing methods for bacterial

microbiota profiling in catfish aquaculture. North American Journal of Aquaculture 00: 1–16. https://doi.org/10.1002/naaq.10309

ONT (2024) Oxford Nanopore Technologies - Oxford Nanopore technology updates show consolidations and updates for a single, high-performance sequencing platform. https://nanoporetech.com/about-us/news/oxford-nanopore-technology-updates-show-consolidations-and-updates-single-high [accessed Jan 25ᵗʰ 2024]

Petrone JR, Rios Glusberger P, George CD, Milletich PL, Ahrens AP, Roesch LFW, Triplett EW (2023) RESCUE: A validated Nanopore pipeline to classify bacteria through long-read, 16S-ITS-23S rRNA sequencing. Frontiers in Microbiology 14: 1201064. https://doi.org/10.3389/fmicb.2023.1201064

Ravi RK, Walton K, Khosroheidari M (2018) MiSeq: A Next Generation Sequencing Platform for Genomic Analysis. Methods in Molecular Biology (Clifton, N.J.) 1706: 223–232. https://doi.org/10.1007/978-1-4939-7471-9_12

Rodríguez-Pérez H, Ciuffreda L, Flores C (2021) NanoCLUST: A species-level analysis of 16S rRNA nanopore sequencing data. Bioinformatics (Oxford, England) 37(11): 1600–1601. https://doi.org/10.1093/bioinformatics/btaa900

Rognes T, Flouri T, Nichols B, Quince C, Mahe F (2016) VSEARCH: A versatile open source tool for metagenomics. PeerJ 4: e2584. https://doi.org/10.7717/peerj.2584

Sanderson ND, Kapel N, Rodger G, Webster H, Lipworth S, Street TL, Peto T, Crook D, Stoesser N (2023) Comparison of R9.4.1/Kit10 and R10/Kit12 Oxford Nanopore flow-cells and chemistries in bacterial genome reconstruction. Microbial Genomics 9(1): mgen000910. https://doi.org/10.1099/mgen.0.000910

Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Research 43(6): e37. https://doi.org/10.1093/nar/gku1341

Schirmer M, D'Amore R, Ijaz UZ, Hall N, Quince C (2016) Illumina error profiles: Resolving fine-scale variation in metagenomic sequencing data. BMC Bioinformatics 17(1): 125. https://doi.org/10.1186/s12859-016-0976-y

SEPA (2023) Scottish Environment Protection Agency: Marine finfish farms – Guidance on the use of DNA in monitoring compliance with seabed mixing zone area limits. https://www.sepa.org.uk/media/594636/guidance-on-the-use-of-dna-in-monitoring-compliance-with-seabed-mixing-zone-area-limits.pdf

Stevens BM, Creed TB, Reardon CL, Manter DK (2023) Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil. Scientific Reports 13(1): 9323. https://doi.org/10.1038/s41598-023-36101-8

Stoeck T, Frühe L, Forster D, Cordier T, Martins CIM, Pawlowski J (2018) Environmental DNA metabarcoding of benthic bacterial communities indicates the benthic footprint of salmon aquaculture. Marine Pollution Bulletin 127: 139–149. https://doi.org/10.1016/j.marpolbul.2017.11.065

Stoler N, Nekrutenko A (2021) Sequencing error profiles of Illumina sequencing instruments. NAR Genomics and Bioinformatics 3(1): lqab019. https://doi.org/10.1093/nargab/lqab019

Tandon D, Dong Y, Hapfelmeier S (2023) Pipeline for species-resolved full-length16S rRNA amplicon nanopore sequencing analysis of low-complexity bacterial microbiota. bioRxiv 2023.2012.2005.570138. https://doi.org/10.1101/2023.12.05.570138

Wang Y, Zhao Y, Bollas A, Wang Y, Au KF (2021) Nanopore sequencing technology, bioinformatics and applications. Nature Biotechnology 39(11): 1348–1365. https://doi.org/10.1038/s41587-021-01108-x

Wick R (2023) Yet another ONT accuracy test: Dorado v0.5.0. Bioinformatics blog. https://doi.org/10.5281/zenodo.10397817

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. 2nd edn. Springer, Dordrecht-Heidelberg-London-New York, 213 pp.

Wilding TA, Stoeck T, Morrissey BJ, Carvalho SF, Coulson MW (2023) Maximising signal-to-noise ratios in environmental DNA-based monitoring. The Science of the Total Environment 858(Pt 1): 159735. https://doi.org/10.1016/j.scitotenv.2022.159735

Xue Z, Kable ME, Marco ML (2018) Impact of DNA sequencing and analysis methods on 16S rRNA gene bBacterial community analysis of dairy products. MSphere 3(5): 00410−00418. https://doi.org/10.1128/mSphere.00410-18

Zhang T, Li H, Ma S, Cao J, Liao H, Huang Q, Chen W (2023) The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. Applied and Environmental Microbiology 89(10): e0060523. https://doi.org/10.1128/aem.00605-23

Zorz J, Li C, Chakraborty A, Gittins DA, Surcon T, Morrison N, Bennett R, MacDonald A, Hubert CRJ (2023) SituSeq: An offline protocol for rapid and remote Nanopore 16S rRNA amplicon sequence analysis. ISME Communications 3(1): 33. https://doi.org/10.1038/s43705-023-00239-3

Supplementary material 1

Incidence-based Jaccard diversity index

Authors: Thorsten Stoeck, Sven Nicolai Katzenmeier, Hans-Werner Breiner, Verena Rubel
Data type: pptx
Explanation note: Incidence-based Jaccard diversity index for all samples included in the study, based on OTUs (Nanopore) and ASVs (Illumina, as well as on taxonomic ranks bacterial family, genus and species (the latter for Nanopore data only).
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.8.121817.suppl1